

撰写要求、模板及评分标准

随着数智时代的到来,文本分析与挖掘技术在许多领域扮演着越来越重要的角色。本课程旨在通过基于 Java 语言的实践操作和理论学习,让大家掌握文本数据分析的基础技能,并应用这些技能来解决实际问题。现在,我们将布置一项结课报告作业,希望同学们能够将所学知识运用到具体的文本分析任务中,展现自己的分析能力和创新思维。

一、选题方向

同学们可选的题目方向包括但不限于以下几个方向:

- 社交媒体数据分析
- 新闻文本分析
- 学术文献内容分析
- 产品评论数据分析
-

所处理数据的语种中英均可,同学们可根据自身情况进行选择。无论哪个语种,所涉及文档的数量不少于 5000 篇。

二、报告要求

报告内容应包括但不限于以下部分:

- 目录结构
- 引言: 介绍选题背景、意义等,也可介绍选定的数据集和分析目标,说明分析该数据集的重要性。
- 数据预处理: 简述数据的收集、清洗和预处理过程。
- 分析方法: 描述所采用分析方法的主要工作原理,并解释为何选择这些方法。分析方法不限课堂上介绍的方法,同学们可以适当扩展。
- 结果与讨论: 以图表的形式展示分析结果,并对分析结果进行深入解读。
- 结论: 基于分析结果,总结研究发现并提出可能的改进方案或后续研究方向。
- 参考文献: 主要参考文献。
- 附件: 正文部分不允许包括代码,不过伪代码可以,所有可直接运行的代码放到附件里,同时可以附一份执行说明文档。对于所涉及的编程语言,大数据管理与应用专业的同学必须采用 Java 语言,信息管理与信息系统(第二学士学位)专业的同学不限。

三、格式与提交

- ◆ 字数：5000 以上（不包括首页、目录、参考文献和附件部分）。
- ◆ 参考文献：至少列出 10 篇与报告主题相关的参考文献，其中至少 6 篇为重要学术期刊论文。请按照学术规范引用参考文献。
- ◆ 格式：请遵守课程提供的报告格式要求。
- ◆ 大模型：不得采用大模型生成报告，可以借助大模型生成分析代码，但必须通过 IDE 调试完成。
- ◆ 提交时间：截止日期为 2025 年 11 月 11 日。请在截止日期前将报告的电子版和纸质版交给各班课代表，课代表收齐后统一给我。

期待大家能够全面展示自己的分析技能和创新思维。如在报告撰写过程中遇到任何问题，欢迎随时咨询。

四、报告评分标准

为确保评审的公平性和严谨性，将根据以下标准对同学位的结课报告进行评分：

（1）内容完整性（40%）

- ✓ 是否对选题背景做了充分的论述
- ✓ 是否对选定数据集进行了充分的描述。
- ✓ 是否详细说明了数据预处理和分析方法。
- ✓ 是否对分析结果进行了深入的解读和讨论。
- ✓ 所提交代码是否可以成功运行。

（2）分析技术应用（30%）

- ✓ 分析方法的选择是否恰当。
- ✓ 是否有效运用了课程中学习的文本分析方法和技术。
- ✓ 结果展示是否清晰、适当。

（3）创新性与逻辑性（20%）

- ✓ 分析过程中是否有创新的思考和方法。
- ✓ 论述是否逻辑清晰，条理分明。

（4）写作质量与学术规范（10%）

- ✓ 语言表达是否流畅，是否有错别字。
- ✓ 报告格式是否符合要求，参考文献是否规范。

请在撰写报告时参照这些评分标准，确保你的报告内容丰富、分析方法应用得当、逻辑清晰，并符合学术规范。注意：如果结课报告由大模型自动生成，最终成绩将不及格。